

# BOUNDS ON THE NUMBER OF INTEGER POINTS IN A POLYTOPE VIA CONCENTRATION ESTIMATES

AUSTIN SHAPIRO

**ABSTRACT.** It is generally hard to count, or even estimate, how many integer points lie in a polytope  $P$ . Barvinok and Hartigan have approached the problem by way of information theory, showing how to efficiently compute a random vector which samples the integer points of  $P$  with (computable) constant mass, but which may also land outside  $P$ . Thus, to count the integer points of  $P$ , it suffices to determine the frequency with which the random vector falls in  $P$ .

We prove a collection of efficiently computable upper bounds on this frequency. We also show that if  $P$  is suitably presented by  $n$  linear inequalities and  $m$  linear equations ( $m$  fixed), then under mild conditions separating the expected value of the above random vector from the origin, the frequency with which it falls in  $P$  is  $O(n^{-m/2})$  as  $n \rightarrow \infty$ . As in the classical *Littlewood–Offord problem*, all results in the paper are obtained by bounding the point concentration of a sum of independent random variables; we sketch connections to previous work on the subject.

## 1. INTRODUCTION

The problem of counting integer points in polytopes has been extensively studied, and appears to be quite difficult in general. It is NP-hard to determine whether an arbitrary integral polytope with  $n$  facets contains an integer point at all [10]. Given this state of affairs, attention has largely shifted to approximating or bounding the number of integer points in a polytope, and the closely related problem of sampling almost uniformly from the set of integer points in a polytope.

For certain classes of polytopes, almost uniform sampling has been achieved by specially constructed Markov chains with good mixing properties. One notable success of this method is due to Jerrum, Sinclair, and Vigoda, who in [13] construct a fully polynomial randomized approximation scheme for the permanent of a 0-1 matrix (equal to the number of integer points in a perfect matching polytope). However, for a general polytope  $P$ , it is not known how to efficiently generate Markov chains which sample almost uniformly from the integer points in  $P$ . A survey of this and other approaches to the problem can be found in [5], [6].

---

*Date:* October 2010.

The author was supported in part by NSF grants DMS 0400617 and DMS 0856640.

In [1], Barvinok and Hartigan proposed a new approach to the problem using the principle of maximum entropy. Given a polytope  $P \subset \mathbb{R}^n$  defined by the inequalities

$$x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0, \quad A\mathbf{x} = \mathbf{b},$$

where  $A \in \mathbb{R}^{m \times n}$  and  $\mathbf{b} \in \mathbb{R}^m$ , they introduce a random vector  $X = (X_1, X_2, \dots, X_n)$  of maximum entropy, subject to the constraints that all coordinates are distributed on  $\mathbb{Z}_{\geq 0}$  and that  $\mathbf{E}[AX] = \mathbf{b}$  (i.e., the mean of  $X$  lies in  $P$ ). This random vector has constant mass  $e^{-H(X)}$  on all points of  $P \cap \mathbb{Z}^n$ , where  $H(X)$  is the *entropy* of the random vector, defined by

$$H(X) := - \sum_{k_1, \dots, k_n \geq 0} \mathbf{Pr}[X = (k_1, \dots, k_n)] \ln \mathbf{Pr}[X = (k_1, \dots, k_n)].$$

Thus  $X$  is, in a sense, a good approximation of the uniform distribution on  $P \cap \mathbb{Z}^n$ . However, not all of the mass of  $X$  lies in  $P$ ; thus we have

$$|P \cap \mathbb{Z}^n| = e^{H(X)} \mathbf{Pr}[X \in P].$$

As it turns out [1], the coordinates of  $X$  are independent and *geometrically distributed*, that is, there exist  $q_j \in [0, 1)$ ,  $1 \leq j \leq n$  so that

$$\mathbf{Pr}[X_j = k] = (1 - q_j)q_j^k \quad \text{for } k \in \mathbb{Z}_{\geq 0}.$$

After a change of parameter  $z_j := \mathbf{E}[X_j] = \frac{q_j}{1 - q_j}$ , the entropy  $H(X)$  may be written as

$$(1) \quad H(X) = \sum_{j=1}^n (z_j + 1) \ln(z_j + 1) - z_j \ln z_j.$$

This is a strictly concave function of  $z_1, \dots, z_n$ , so it can be maximized efficiently by (e.g.) interior point methods (for details, see [1]). Thus the parameters  $q_j$ , and with them the distribution and entropy of  $X$ , are efficiently computable. Hence, the outstanding question is how to bound the factor  $\mathbf{Pr}[X \in P]$ , particularly under weak assumptions (i.e., when a local central limit theorem is not feasible). This paper offers several upper bounds.

## 2. SUMMARY OF RESULTS

**2.1. Definitions and notation.** Throughout this paper,  $A$  always denotes an  $m \times n$  matrix with real entries; we assume that  $n > m$  and that  $\text{rank}(A) = m$ . We denote the columns of  $A$  by  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ . The random vector  $X = (X_1, X_2, \dots, X_n)$  is defined as in the introduction, so as to maximize the entropy  $H(X)$  subject to the constraint  $\mathbf{E}[AX] = \mathbf{b} = (b_1, b_2, \dots, b_m) \in \mathbb{R}^m$ . We define the parameters  $q_j, z_j$  as in the introduction.

We define the *point concentration* of a discrete random variable  $Y$  by

$$\text{conc}(Y) := \max_y \mathbf{Pr}[Y = y].$$

An upper bound on  $\text{conc}(AX)$  is, necessarily, also an upper bound on  $\Pr[AX = \mathbf{b}] = \Pr[X \in P]$ . Therefore, we have

$$(2) \quad |P \cap \mathbb{Z}^n| \leq e^{H(X)} \text{conc}(AX).$$

**2.2. Results.** Under the hypotheses above, we prove:

**Theorem 1.**

$$\begin{aligned} |P \cap \mathbb{Z}^n| &\leq e^{H(X)} \min_{\mathbf{a}_{j_1}, \dots, \mathbf{a}_{j_m} \text{ lin. indep.}} (1 - q_{j_1})(1 - q_{j_2}) \cdots (1 - q_{j_m}) \\ &= e^{H(X)} \min_{\mathbf{a}_{j_1}, \dots, \mathbf{a}_{j_m} \text{ lin. indep.}} \prod_{i=1}^m \frac{1}{z_{j_i} + 1}. \end{aligned}$$

**Corollary 1.** Let  $I_1, I_2, \dots, I_p$  be  $m$ -element subsets of  $\{1, 2, \dots, n\}$ ,

$$I_k = \{j_{k1}, j_{k2}, \dots, j_{km}\},$$

such that  $\mathbf{a}_{j_{k1}}, \dots, \mathbf{a}_{j_{km}}$  form a basis for  $\mathbb{R}^m$  ( $1 \leq k \leq p$ ), and such that  $I_1 \cup I_2 \cup \dots \cup I_p = \{1, 2, \dots, n\}$ . Then

$$|P \cap \mathbb{Z}^n| \leq e^{H(X)} \left( \frac{1}{\mathbf{E}[\bar{X}] + 1} \right)^m,$$

where  $\bar{X}$  is a geometrically distributed random variable with entropy equal to  $\frac{1}{pm}H(X)$ .

(A formula for the entropy of a geometrically distributed random variable is given in section 1, (1).)

**Theorem 2.** Suppose that  $n = pm$  for some integer  $p$ , that  $A$  has integer entries, and that  $\mathbf{a}_{(k-1)m+1}, \mathbf{a}_{(k-1)m+2}, \dots, \mathbf{a}_{km}$  are linearly independent for  $1 \leq k \leq p$ . Assume that  $\langle \mathbf{a}_j, \mathbf{b} \rangle > 0$  for  $1 \leq j \leq n$ . Define

$$q_i^\vee := \min\{q_{(k-1)m+i} : 1 \leq k \leq p\} \quad (1 \leq i \leq m).$$

Then there exist constants  $C = C(q_1^\vee, \dots, q_m^\vee)$  and  $C' = C'(q_1^\vee, \dots, q_m^\vee)$ , with  $C' < 1$ , such that

$$|P \cap \mathbb{Z}^n| \leq e^{H(X)} (Cp^{-m/2} + (C')^p).$$

(In fact, there is a one-parameter family of pairs of constants  $(C, C')$  for which this statement holds. Explicit formulas and bounds for  $C$  and  $C'$  are provided in section 5.)

**Theorem 3.** Suppose that  $n = pm$  for some integer  $p$  and that, for each  $i = 1, 2, \dots, m$ , we have  $\mathbf{a}_i = \mathbf{a}_{m+i} = \mathbf{a}_{2m+i} = \dots = \mathbf{a}_{(p-1)m+i}$ , where  $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$  is a basis for  $\mathbb{R}^m$ . (That is to say, the columns of  $A$  cycle through a basis of  $\mathbb{R}^m$  periodically.) Then

$$|P \cap \mathbb{Z}^n| \lesssim e^{H(X)} \prod_{i=1}^m \left( \frac{\pi p}{6} (\lfloor \mathbf{E}(X_i) + 1 \rfloor^2 - 1) \right)^{-1/2}.$$

(Here  $\lesssim$  means that, given fixed  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ , the expression on the left side is bounded above by a function which is asymptotic to the expression on the right side as  $p \rightarrow \infty$ .)

**2.3. Plan of paper.** In section 3, we discuss these results in the context of prior work, and give examples of their use. In section 4, we prove Theorem 1 and Corollary 1. The most substantial portion of the paper is section 5, in which we prove Theorem 2, then bound the constants appearing in it. In section 6, we prove Theorem 3.

### 3. DISCUSSION AND EXAMPLES

The concentration of sums of random variables is a richly studied subject. The particular program of obtaining *upper* bounds, sometimes called “anti-concentration results,” may be considered to have originated with the *Littlewood–Offord problem*. This problem asked for the maximum concentration of

$$\varepsilon_1 a_1 + \varepsilon_2 a_2 + \dots + \varepsilon_n a_n$$

when  $a_1, a_2, \dots, a_n$  are integers and  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  are symmetric Bernoulli random variables. The exact solution, which is of order  $O(n^{-1/2})$ , was provided by Erdős [9].

Halász [11] extended this result to random sums of  $m$ -vectors (again with symmetric Bernoulli coefficients), obtaining a bound of order  $O(n^{-m/2})$  under conditions ensuring that the vectors are reasonably “spread out” in  $\mathbb{R}^m$  (i.e., not excessively close to a proper subspace). Halász’s results pertain to the *small ball* concentration of  $\varepsilon_1 a_1 + \varepsilon_2 a_2 + \dots + \varepsilon_n a_n$ , but can be specialized to point concentration. These results, which Halász proved using a Fourier-theoretic lemma of Esséen, were subsequently reproduced by Oskolkov [12, notes by Howard] using *rearrangement inequalities*. Theorem 2, herein, arrives at a similar conclusion when the Bernoulli coefficients are replaced by geometric ones. In particular, Theorem 2 implies the following Gaussian-like asymptotics:

**Corollary 2.** *Suppose that a subset of the columns of  $A$  can be partitioned into  $p$  bases for  $\mathbb{R}^m$ . Then for  $\min_j q_j$  bounded away from 0, the point concentration of  $AX$  is  $O(p^{-m/2})$  as  $p \rightarrow \infty$ .*

Our proof of Theorem 2 hews closely to the method of [12]. For other approaches to anti-concentration inequalities, see [15], [16].

Theorem 2 is essentially an asymptotic result; although we give explicit formulas for  $C$  and  $C'$ , the bounds obtained from Theorem 2 are typically only strong when  $p$  is large, i.e., when  $n \gg m$ . (For further remarks on this theme, see the end of section 5.1.) By contrast, Theorem 1 and its corollary are non-asymptotic, and are apparently most effective when  $n \not\gg m$ . They are also relatively straightforward, but do not capture the  $O(p^{-m/2})$  behavior of  $\text{conc}(AX)$ . Thus, Theorem 1 and Theorem 2 may be seen as filling somewhat different niches. Theorem 3 gives a

more ideal bound, combining all the attractive features of Theorems 1 and 2, but is pertinent only to a very special case (the easiest, where a local central limit theorem is available). Relying for its proof on notions from the theory of partially ordered sets, Theorem 3 may serve as a suggestion of how combinatorics can be brought to bear on this problem.

**3.1. Examples.** Given nonnegative vectors  $R \in \mathbb{R}^r$ ,  $C \in \mathbb{R}^s$ , the *transportation polytope*  $\Pi(R, C)$  is defined as the set of all nonnegative  $r \times s$  matrices whose row sums and column sums are the coordinates of  $R$  and  $C$ , respectively. Such a matrix with integer entries is called a *contingency table*.

We may use Theorem 1 to bound the number of  $4 \times 4$  contingency tables with given “margins”  $R$  and  $C$ . For example, let  $R = (108, 286, 71, 127)$  and  $C = (220, 215, 93, 64)$ , as in a table appearing in [7] which has become a standard benchmark in the literature on contingency tables. The actual number of tables with these row and column sums is  $1.23 \times 10^{15}$ . (It can be computed exactly, as the dimension is fairly low: the defining matrix  $A$  for  $\Pi(R, C)$  is  $7 \times 16$ .)

Let  $X$  be a random matrix taking the maximum-entropy distribution on  $\mathbb{Z}_{\geq 0}^{4 \times 4}$ , under the constraint that  $\mathbf{E}[X] \in \Pi(R, C)$ . Solving the convex optimization problem described in section 1, we compute

$$\mathbf{E}[X] = \begin{pmatrix} 36.4 & 36.0 & 20.6 & 14.9 \\ 117.2 & 113.4 & 34.3 & 21.2 \\ 22.2 & 22.0 & 15.1 & 11.7 \\ 44.2 & 43.6 & 23.0 & 16.2 \end{pmatrix}$$

and  $H(X) = 2.96 \times 10^{30}$ . Theorem 1 then gives

$$\begin{aligned} |P \cap \mathbb{Z}^n| &\leq \frac{2.96 \times 10^{30}}{(1 + 36.4)(1 + 117.2)(1 + 113.4)(1 + 34.3)(1 + 21.2)(1 + 22.2)(1 + 44.2)} \\ &= 7.14 \times 10^{18}, \end{aligned}$$

off by a factor of about 5800. Computation of similar examples suggests that the relative error depends mainly on the dimensions of  $R$  and  $C$ , and not on the magnitude of their entries.

Theorem 2 performs relatively poorly in these examples, but is much more effective than Theorem 1 when  $n$  is large compared to  $m$ . For instance, consider the simplex

$$\Sigma^n(r) := \{(x_1, \dots, x_n) : x_1, \dots, x_n \geq 0, \quad \|\mathbf{x}\|_1 = r\},$$

which has  $\binom{n+r-1}{r}$  integer points. Let  $0 < \delta < \frac{1}{2}$ . Then, choosing  $\gamma = \frac{\pi r^\delta}{\sqrt{n}}$  in the statement of Theorem 2a (see section 5), one obtains as a conclusion an upper bound on  $|\Sigma^n(r) \cap \mathbb{Z}^n|$  which is precisely asymptotic to  $\binom{n+r-1}{r}$  as  $n \rightarrow \infty$ , if  $r$  grows as  $\Theta(n^\varepsilon)$  for some  $\varepsilon \in (0, 1)$ .

For  $r = 10$  and  $n = 1000$ , the optimal result of Theorem 2a (achieved when  $\gamma = 0.172$ ) is an upper bound of  $3.14 \times 10^{23}$ , which may be compared with an exact count of  $2.88 \times 10^{23}$  integer points. By comparison, when  $r = 100$  and  $n = 10000$ , the optimal result of Theorem 2a (achieved when  $\gamma = 0.0645$ ) is an upper bound of  $1.774 \times 10^{242}$  integer points; the exact count is  $1.755 \times 10^{242}$ , and the relative error is about 1.1%.

#### 4. PROOFS OF THEOREM 1 AND COROLLARY 1

We prove Theorem 1 by means of the following simple fact:

**Lemma 1.** *If  $X, Y$  are independent, discrete random variables, then  $\text{conc}(X+Y) \leq \text{conc}(X)$ .*

*Proof.* Observe that  $\text{conc}(X+Y)$  is a weighted average of values of the probability mass function of  $X$ , of which the largest is  $\text{conc}(X)$ .  $\square$

*Proof of Theorem 1.* Using Lemma 1 and the previously mentioned properties of geometric random variables,

$$\begin{aligned} \text{conc}(X_1 \mathbf{a}_1 + \cdots + X_n \mathbf{a}_n) &\leq \min_{\mathbf{a}_{j_1}, \dots, \mathbf{a}_{j_m} \text{ lin. indep.}} \text{conc}(X_{j_1} \mathbf{a}_{j_1} + \cdots + X_{j_m} \mathbf{a}_{j_m}) \\ &\leq \min_{\mathbf{a}_{j_1}, \dots, \mathbf{a}_{j_m} \text{ lin. indep.}} \Pr[X_{j_1} = \cdots = X_{j_m} = 0] \\ &= \min_{\mathbf{a}_{j_1}, \dots, \mathbf{a}_{j_m} \text{ lin. indep.}} (1 - q_{j_1})(1 - q_{j_2}) \cdots (1 - q_{j_m}) \\ &= \min_{\mathbf{a}_{j_1}, \dots, \mathbf{a}_{j_m} \text{ lin. indep.}} \prod_{i=1}^m \frac{1}{z_{j_i} + 1}. \end{aligned}$$

By section 2.1, (2), it follows that

$$\begin{aligned} |P \cap \mathbb{Z}^n| &\leq e^{H(X)} \min_{\mathbf{a}_{j_1}, \dots, \mathbf{a}_{j_m} \text{ lin. indep.}} (1 - q_{j_1})(1 - q_{j_2}) \cdots (1 - q_{j_m}) \\ &= e^{H(X)} \min_{\mathbf{a}_{j_1}, \dots, \mathbf{a}_{j_m} \text{ lin. indep.}} \prod_{i=1}^m \frac{1}{z_{j_i} + 1}. \quad \blacksquare \end{aligned}$$

To prove Corollary 1, we will require this fact whose proof is deferred until after the proof of Corollary 1:

**Lemma 2.** *Among all vectors  $Y := (Y_1, Y_2, \dots, Y_m)$  of independent, geometrically distributed random variables with fixed joint entropy  $\Omega$ , the highest concentration  $\text{conc}(Y)$  is achieved when  $Y_1, Y_2, \dots, Y_m$  are identically distributed.*

*Proof of Corollary 1.* For  $I \subset \{1, 2, \dots, n\}$ , let  $H(X_I)$  denote the joint entropy of  $\{X_j : j \in I\}$ . Since  $X_1, \dots, X_n$  are pairwise independent, we have  $H(X_I) = \sum_{j \in I} H(X_j)$ .

Since the sets  $I_1, I_2, \dots, I_p$  cover  $\{1, 2, \dots, n\}$ , we have

$$H(X) \leq \sum_{k=1}^p H(X_{I_k}),$$

and thus by the pigeonhole principle

$$H(X_{I_k}) \geq \frac{1}{p} H(X)$$

for some  $k \in \{1, \dots, p\}$ . By Lemma 2, the concentration of the vector  $(X_{j_{k1}}, \dots, X_{j_{km}})$  is maximized when  $X_{j_{k1}}, \dots, X_{j_{km}}$  are identically distributed. In this case, each has entropy equal to  $\frac{1}{m} H(X_{I_k})$ , which is greater than or equal to  $H(\bar{X}) = \frac{1}{pm} H(X)$ ; we pause to note that the entropy and the expectation of a geometric random variable are monotonically increasing functions of one another. Thus (as in the proof of Theorem 1),

$$\begin{aligned} \text{conc}(AX) &\leq \text{conc}(X_{j_{k1}} \mathbf{a}_{j_{k1}} + \dots + X_{j_{km}} \mathbf{a}_{j_{km}}) \\ &\leq \left( \frac{1}{\mathbf{E}[\bar{X}] + 1} \right)^m, \end{aligned}$$

so Corollary 1 follows by section 2.1, (2). ■

*Proof of Lemma 2.* Since  $Y_i$  is geometrically distributed ( $1 \leq i \leq m$ ), there exist parameters  $r_i \in [0, 1)$  such that

$$\Pr[Y_i = k] = (1 - r_i) r_i^k \quad \text{for } k \in \mathbb{Z}_{\geq 0}.$$

The concentration of  $Y$  is  $\prod_{i=1}^m (1 - r_i)$ , so we must show that this expression is maximized (for fixed  $\Omega$ ) when  $r_1 = \dots = r_m$ .

We introduce the changes of variable  $s_i := \frac{1}{1-r_i}$ ,  $t_i := \ln s_i$ . (Thus  $1 - r_i = \frac{1}{s_i}$ , and  $s_i = e^{t_i}$ , where  $t_i \in [0, \infty)$ .) Also, let

$$\omega(t) := (1 - e^t) \ln(1 - e^{-t}) + t.$$

Now

$$\begin{aligned}
\Omega &= \sum_{i=1}^m \frac{r_i}{1-r_i} \ln \frac{1}{r_i} + \ln \frac{1}{1-r_i} \\
&= \sum_{i=1}^m (s_i - 1) \ln \frac{s_i}{s_i - 1} + \ln s_i \\
&= \sum_{i=1}^m (e^{t_i} - 1) \ln \frac{e^{t_i}}{e^{t_i} - 1} + t_i \\
&= \sum_{i=1}^m (1 - e^{-t_i}) \ln(1 - e^{-t_i}) + t_i \\
&= \sum_{i=1}^m \omega(t_i),
\end{aligned}$$

and

$$\prod_{i=1}^m (1 - r_i) = \exp \left( - \sum_{i=1}^m t_i \right).$$

The following three statements are equivalent:

- (1) For  $\Omega$  fixed,  $\prod_i (1 - r_i)$  is maximized when  $r_1 = \dots = r_m$ .
- (2) For  $\Omega$  fixed,  $\sum_i t_i$  is minimized when  $t_1 = \dots = t_m$ .
- (3) If  $\sum_i t_i$  is fixed and  $\Omega$  free to vary, then  $\Omega$  is maximized when  $t_1 = \dots = t_m$ .

The equivalence of statements (1) and (2) is clear. To see that (2) and (3) are equivalent, it is enough to observe that  $\Omega$  is increasing with respect to each of  $t_1, \dots, t_m$ . Thus to prove (1), which is the assertion of the lemma, it will suffice for us to prove (3).

Writing  $s := e^t$ , we obtain

$$\begin{aligned}
\frac{d\omega}{dt} &= (1 - e^t) \left( \frac{e^{-t}}{1 - e^{-t}} \right) - e^t \ln(1 - e^{-t}) + 1 \\
&= -e^t \ln(1 - e^{-t})
\end{aligned}$$



and

$$\begin{aligned}
\frac{d^2\omega}{dt^2} &= -e^t \cdot \frac{e^{-t}}{1 - e^{-t}} - e^t \ln(1 - e^{-t}) \\
&= -\frac{1}{1 - \frac{1}{s}} - s \ln\left(1 - \frac{1}{s}\right) \\
&= -\frac{s}{s-1} + s \ln \frac{s}{s-1} \\
&= -s \left(\frac{1}{s-1}\right) + s \ln\left(1 + \frac{1}{s-1}\right) \\
&\leq 0,
\end{aligned}$$

since  $\ln(1+x) \leq x$  for  $x \geq 0$ . This shows that  $\omega(t)$  is concave for  $t \geq 0$ , which implies (3) and so completes the proof of the lemma. ■

## 5. PROOF OF THEOREM 2

We begin by restating the theorem with explicit formulas for all constants:

**Theorem 2a.** *Assume the definitions and notation from section 2.1.*

*Suppose that  $n = pm$  for some integer  $p$ , that  $A$  has integer entries, and that  $\mathbf{a}_{(k-1)m+1}, \mathbf{a}_{(k-1)m+2}, \dots, \mathbf{a}_{km}$  are linearly independent for  $1 \leq k \leq p$ . Assume that  $\langle \mathbf{a}_j, \mathbf{b} \rangle > 0$  for  $1 \leq j \leq n$ . Let  $\gamma > 0$ . Define constants*

$$\begin{aligned}
\alpha_j &:= \frac{2q_j}{(1 - q_j)^2} & (1 \leq j \leq n), \\
\alpha_i^\vee &:= \min\{\alpha_{(k-1)m+i} : 1 \leq k \leq p\} & (1 \leq i \leq m), \\
q_i^\vee &:= \min\{q_{(k-1)m+i} : 1 \leq k \leq p\} & (1 \leq i \leq m), \\
c_i &:= \max \left\{ \frac{1}{\gamma^2} \ln \left[ 1 + \alpha_i^\vee \left( 1 - \cos \frac{\gamma}{\sqrt{\alpha_i^\vee}} \right) \right], \quad \frac{1}{\alpha_i^\vee \pi^2} \ln \left[ 1 + 2\alpha_i^\vee \right] \right\} & (1 \leq i \leq m), \\
C &:= \prod_{i=1}^m (2\pi c_i \alpha_i^\vee)^{-1/2}, \\
C' &:= \max_{1 \leq i \leq m} e^{-\gamma^2 c_i / 2}.
\end{aligned}$$

Then

$$|P \cap \mathbb{Z}^n| \leq e^{H(X)} (Cp^{-m/2} + (C')^p).$$

All notation introduced in Theorem 2a is used throughout this section, and all its hypotheses (importantly, the integrality of  $A$ ) are assumed to hold. In subsection 5.1, we introduce a series of definitions and lemmas, then prove Theorem 2a under assumption of the lemmas. In subsection 5.2, we prove the lemmas in turn. For bounds on the constants  $C$  and  $C'$ , see subsection 5.3.

### 5.1. Supporting results and proof of Theorem 2a.

**Definition 1.** For  $1 \leq k \leq p$ , define the function  $\Pi_k : \mathbb{R}^m \rightarrow \mathbb{R}$  by

$$\begin{aligned} \Pi_k(\mathbf{t}) &:= \prod_{j=(k-1)m+1}^{km} \frac{1}{\sqrt{1 + \alpha_j(1 - \cos\langle \mathbf{t}, \mathbf{a}_j \rangle)}} & \text{for } \mathbf{t} \in (-\pi, \pi]^m, \\ \Pi_k(\mathbf{t}) &:= 0 & \text{for } \mathbf{t} \notin (-\pi, \pi]^m. \end{aligned}$$

**Lemma 3.** Given the definition above,

$$\text{conc}(AX) \leq \frac{1}{(2\pi)^m} \int_{(-\pi, \pi]^m} \Pi_1 \Pi_2 \cdots \Pi_p \, d\mathbf{t}.$$

**Definition 2.** Given a measurable function  $\Phi : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$ , we define its epigraphs

$$\Gamma_{\geq \tau}(\Phi) := \{\mathbf{t} \in \mathbb{R}^m : \Phi(\mathbf{t}) \geq \tau\}$$

for all  $\tau > 0$ .

Suppose  $\Phi$  vanishes at infinity, meaning that  $\Gamma_{\geq \tau}(\Phi)$  has finite volume for each  $\tau > 0$ . Then we define its symmetrically decreasing rearrangement as the function  $\Phi^* : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$  given by

$$\Phi^*(\mathbf{t}) := \max \left\{ \tau : \text{vol}(\Gamma_{\geq \tau}(\Phi)) \geq \|\mathbf{t}\|^m v_m \right\},$$

where  $v_m$  denotes the volume of the unit ball in  $\mathbb{R}^m$ .

The theory of symmetrically decreasing rearrangements is treated in [4], and we do not develop it fully here. The important properties of  $\Phi^*$  are that

- $\Phi^*$  is symmetrically decreasing, i.e.,  $\|\mathbf{t}\| \geq \|\mathbf{s}\| \Rightarrow \Phi^*(\mathbf{t}) \leq \Phi^*(\mathbf{s})$ ; and
- $\Phi^*$  is equimeasurable with  $\Phi$ , i.e.,  $\text{vol}(\Gamma_{\geq \tau}(\Phi^*)) = \text{vol}(\Gamma_{\geq \tau}(\Phi))$  for all  $\tau > 0$ .

Note that  $\Phi^*$  is the unique function with these properties, up to a difference on a set of measure zero.

**Lemma 4.** Given the definition above,

$$\int_{(-\pi, \pi]^m} \Pi_1 \Pi_2 \cdots \Pi_p \, d\mathbf{t} \leq \int_{\mathbb{R}^m} \Pi_1^* \Pi_2^* \cdots \Pi_p^* \, d\mathbf{t}.$$

**Definition 3.** For  $1 \leq k \leq p$ , define the function  $\Pi_k^{\text{rect}} : \mathbb{R}^m \rightarrow \mathbb{R}$  by

$$\begin{aligned} \Pi_k^{\text{rect}}(\mathbf{t}) &:= \prod_{i=1}^m \frac{1}{\sqrt{1 + \alpha_{(k-1)m+i}(1 - \cos \mathbf{t}_i)}} & \text{for } \mathbf{t} \in (-\pi, \pi]^m, \\ \Pi_k^{\text{rect}}(\mathbf{t}) &:= 0 & \text{for } \mathbf{t} \notin (-\pi, \pi]^m. \end{aligned}$$

The formula for  $\Pi_k^{\text{rect}}$  differs from that for  $\Pi_k$  in that the linear form  $\langle \mathbf{t}, \mathbf{a}_{(k-1)m+i} \rangle$  in the denominator of  $\Pi_k$  is replaced by  $\mathbf{t}_i$ . Effectively, each basis

$$\mathbf{a}_{(k-1)m+1}, \mathbf{a}_{(k-1)m+2}, \dots, \mathbf{a}_{km}$$

of  $\mathbb{R}^m$  is replaced by a standard basis. This will make  $\Pi_k^{\text{rect}}$  easier to work with than  $\Pi_k$ .

**Lemma 5.** *Let  $1 \leq k \leq p$ . Then*

$$\text{vol}(\Gamma_{\geq \tau}(\Pi_k^{\text{rect}})) = \text{vol}(\Gamma_{\geq \tau}(\Pi_k))$$

for all  $\tau > 0$ , and  $(\Pi_k^{\text{rect}})^* \equiv \Pi_k^*$ .

**Lemma 6** (Isotonicity of rearrangement). *Suppose  $\Phi, \Psi : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$  are measurable functions vanishing at infinity. Let  $\tau$  denote a constant. Then:*

- (1) *If  $\Phi(\mathbf{t}) \geq \Psi(\mathbf{t})$  for all  $\mathbf{t}$ , then  $\Phi^*(\mathbf{t}) \geq \Psi^*(\mathbf{t})$  for all  $\mathbf{t}$ .*
- (2) *If  $\Phi(\mathbf{t}) \geq \max\{\Psi(\mathbf{t}), \tau\}$  for all  $\mathbf{t}$ , then  $\Phi^*(\mathbf{t}) \geq \max\{\Psi^*(\mathbf{t}), \tau\}$  for all  $\mathbf{t}$ .*

**Lemma 7.** *Define  $\alpha_i^\vee$  and  $c_i$  as in the statement of Theorem 2a.*

*Then, for  $0 \leq t \leq \min\left\{\frac{\gamma}{\sqrt{\alpha_i^\vee}}, \pi\right\}$ , we have  $1 + \alpha_i^\vee(1 - \cos t) \geq e^{c_i \alpha_i^\vee t^2}$ .*

**Lemma 8.** *For each  $k = 1, 2, \dots, p$ , and for all  $\mathbf{t} \in \mathbb{R}^m$ , we have*

$$\Pi_k^{\text{rect}}(\mathbf{t}) \leq \max \left\{ \prod_{i=1}^m e^{-c_i \alpha_i^\vee \mathbf{t}_i^2 / 2}, \quad C' \right\}.$$

Given the above lemmas, we can prove Theorem 2a:

*Proof of Theorem 2a.* Using Lemmas 3, 4, and 5, we have

$$\begin{aligned} \text{conc}(AX) &\leq \frac{1}{(2\pi)^m} \int_{(-\pi, \pi]^m} \Pi_1 \Pi_2 \cdots \Pi_p \, d\mathbf{t} \\ &\leq \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} \Pi_1^* \Pi_2^* \cdots \Pi_p^* \, d\mathbf{t} \\ &= \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} (\Pi_1^{\text{rect}})^* (\Pi_2^{\text{rect}})^* \cdots (\Pi_p^{\text{rect}})^* \, d\mathbf{t}. \end{aligned}$$

We may instead take either of the last two integrals over  $B$ , the closed ball of volume  $(2\pi)^m$  centered at the origin in  $\mathbb{R}^m$ , since the integrands are zero outside this ball.

By Lemmas 6 and 8, we have

$$\begin{aligned}
& \frac{1}{(2\pi)^m} \int_B (\Pi_1^{\text{rect}})^* (\Pi_2^{\text{rect}})^* \cdots (\Pi_p^{\text{rect}})^* d\mathbf{t} \\
& \leq \frac{1}{(2\pi)^m} \int_B \prod_{k=1}^p \left( \max \left\{ \left( \prod_{i=1}^m e^{-c_i \alpha_i^\vee \mathbf{t}_i^2 / 2} \right)^*, C' \right\} \right) d\mathbf{t} \\
& = \frac{1}{(2\pi)^m} \int_B \left( \max \left\{ \left( \prod_{i=1}^m e^{-c_i \alpha_i^\vee \mathbf{t}_i^2 / 2} \right)^*, C' \right\} \right)^p d\mathbf{t} \\
& = \frac{1}{(2\pi)^m} \int_{(-\pi, \pi]} \left( \max \left\{ \prod_{i=1}^m e^{-c_i \alpha_i^\vee \mathbf{t}_i^2 / 2}, C' \right\} \right)^p d\mathbf{t}.
\end{aligned}$$

This last integral is bounded above by

$$\begin{aligned}
& \frac{1}{(2\pi)^m} \left[ \int_{\mathbb{R}^m} \left( \prod_{i=1}^m e^{-c_i \alpha_i^\vee \mathbf{t}_i^2 / 2} \right)^p d\mathbf{t} + \int_B (C')^p d\mathbf{t} \right] \\
& = \frac{1}{(2\pi)^m} \left[ \int_{\mathbb{R}^m} \exp \left( -p \sum_{i=1}^m -c_i \alpha_i^\vee \mathbf{t}_i^2 / 2 \right) d\mathbf{t} + (2\pi)^m (C')^p d\mathbf{t} \right] \\
& = \frac{1}{(2\pi)^m} \cdot (2\pi)^{m/2} p^{-m/2} \prod_{i=1}^m (c_i \alpha_i^\vee)^{-1/2} + (C')^p \\
& = Cp^{-m/2} + (C')^p.
\end{aligned}$$

Now, a technical remark. In integrating the Gaussian term, we assumed  $c_i \alpha_i^\vee > 0$ . To see why this is necessarily true, note that we assumed, in the statement of Theorem 2a, that  $\langle \mathbf{a}_j, \mathbf{b} \rangle > 0$  for  $1 \leq j \leq n$ . Thus  $P$  is not contained in any coordinate hyperplane of  $\mathbb{R}^n$ . Recall section 1, (1), which gives the entropy  $H(X)$  in terms of  $z_1, z_2, \dots, z_n$  (the coordinates of  $\mathbf{E}[X]$ ). One may check that  $\frac{\partial}{\partial z_j} H(X) = \infty$  when  $z_j = 0$ , but is finite when  $z_j > 0$ . Therefore the maximum-entropy distribution for  $X$  does not take expected value on a coordinate hyperplane; therefore,  $c_i \alpha_i^\vee > 0$ .

Theorem 2a now follows by section 2.1, (2). ■

*Remarks.* Our strategy for bounding  $\text{conc}(AX)$ , carried out above, may be motivated as follows. First, we obtain an integral formula for the probability mass function of  $AX$ , derived from its Fourier transform (Lemma 3). The integrand splits into  $n$  factors, which we then group into maximal subproducts such that the factors in each subproduct behave like independent random variables on the domain of integration. The worst case is now that these subproducts themselves are “completely non-independent,” that is, that they decay identically; this is the significance of Lemmas 4 and 5, and of the definitions of  $q_i^\vee$  and  $\alpha_i^\vee$ . We bound the decay of the

integrand near the origin by a Gaussian (Lemma 8), explaining the appearance of the  $Cp^{-m/2}$  term in the conclusion of Theorem 2a. Away from the origin, we simply bound each subproduct by the constant  $C'$ , giving the  $(C')^p$  term. The parameter  $\gamma$  controls the boundary between the two approximation regimes.

This two-regime bound (with arbitrary parameter  $\gamma$ ) is sufficient for Corollary 2, as the  $(C')^p$  term is asymptotically negligible as  $p \rightarrow \infty$ . However, for non-asymptotic computations, the crudity of the approximation away from the origin is very noticeable. The  $(C')^p$  term can be replaced by a more sensitive approximation, at the cost of simplicity. We do not pursue this goal here.

## 5.2. Proofs of preceding lemmas.

*Proof of Lemma 3.* In [1], Lemma 8.1, the following integral representation is proved:

$$\Pr[AX = \mathbf{b}] = \frac{1}{(2\pi)^m} \int_{(-\pi, \pi]^m} e^{-i\langle \mathbf{t}, \mathbf{b} \rangle} \prod_{j=1}^n \frac{1 - q_j}{1 - q_j e^{i\langle \mathbf{t}, \mathbf{a}_j \rangle}} d\mathbf{t},$$

where  $\mathbf{b}$  is an arbitrary  $\mathbb{Z}_{\geq 0}$ -vector. It follows that

$$\begin{aligned} \text{conc}(AX) &\leq \frac{1}{(2\pi)^m} \int_{(-\pi, \pi]^m} \left| e^{-i\langle \mathbf{t}, \mathbf{b} \rangle} \prod_{j=1}^n \frac{1 - q_j}{1 - q_j e^{i\langle \mathbf{t}, \mathbf{a}_j \rangle}} \right| d\mathbf{t} \\ &= \frac{1}{(2\pi)^m} \int_{(-\pi, \pi]^m} \prod_{j=1}^n \frac{1 - q_j}{\sqrt{1 + q_j^2 - 2q_j \cos \langle \mathbf{t}, \mathbf{a}_j \rangle}} d\mathbf{t} \\ &= \frac{1}{(2\pi)^m} \int_{(-\pi, \pi]^m} \Pi_1 \Pi_2 \cdots \Pi_p d\mathbf{t}, \end{aligned}$$

where the last two steps are straightforward simplification.  $\square$

*Proof of Lemma 4.* The *Hardy–Littlewood inequality* [4] states that for measurable functions  $\Phi, \Psi : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$  vanishing at infinity, one has

$$\int_{\mathbb{R}^m} \Phi(\mathbf{t}) \Psi(\mathbf{t}) d\mathbf{t} \leq \int_{\mathbb{R}^m} \Phi^*(\mathbf{t}) \Psi^*(\mathbf{t}) d\mathbf{t},$$

provided that the integral on the right-hand side converges. Thus we obtain

$$\int_{(-\pi, \pi]^m} \Pi_1 \Pi_2 \cdots \Pi_p d\mathbf{t} \leq \int_{\mathbb{R}^m} \Pi_1^* \Pi_2^* \cdots \Pi_p^* d\mathbf{t}$$

by induction on  $p$ .  $\square$

*Proof of Lemma 5.*

Let  $A_*$  be the  $m \times m$  matrix whose rows are  $\mathbf{a}_{(k-1)m+1}^T, \mathbf{a}_{(k-1)m+2}^T, \dots, \mathbf{a}_{km}^T$ , and define  $\mathcal{A}_* : \mathbb{R}^m \rightarrow \mathbb{R}^m$  as the linear map  $\mathbf{t} \mapsto A_* \mathbf{t}$ . Thus,

$$\mathcal{A}_*(\mathbf{t})_i = \langle \mathbf{t}, \mathbf{a}_{(k-1)m+i} \rangle \quad (1 \leq i \leq m).$$

This map  $\mathcal{A}_*$  scales the volume of measurable sets uniformly by a factor of  $d := |\det(A_*)|$ , and takes the lattice  $\Lambda := (2\pi\mathbb{Z})^m$  to the lattice

$$\Lambda' := 2\pi\mathbb{Z}[\mathbf{col}_1(A_*), \mathbf{col}_2(A_*), \dots, \mathbf{col}_m(A_*)].$$

Let  $K := (-\pi, \pi]^m$  and let  $K' := \mathcal{A}_*(K)$ . Since  $K$  is a fundamental region of  $\Lambda$ , it follows that  $K'$  is a fundamental region of  $\Lambda'$ . Moreover, we assumed  $A$  to have integer entries, so  $\Lambda'$  is a sublattice of index  $d$  in  $\Lambda$ , and the induced map of tori  $\phi : \mathbb{R}^m/\Lambda' \rightarrow \mathbb{R}^m/\Lambda$  is an even covering of order  $d$ .

Identifying  $K$  with  $\mathbb{R}^m/\Lambda$  and  $K'$  with  $\mathbb{R}^m/\Lambda'$ , we may regard  $\phi$  as a map from  $K'$  to  $K$ , and  $\phi \circ \mathcal{A}_*$  as a self-map of  $K$ . If  $U \subseteq K$  is a measurable set, then  $(\phi \circ \mathcal{A}_*)^{-1}(U)$  is the union of  $d$  disjoint preimages each of volume  $\frac{\text{vol}(U)}{d}$ . Thus,  $\text{vol}((\phi \circ \mathcal{A}_*)^{-1}(U)) = \text{vol}(U)$ .

Observe that  $\cos \mathbf{t}_i = \cos(\phi(\mathbf{t})_i)$  for all  $\mathbf{t}$ . Therefore

$$\begin{aligned} \Gamma_{\geq \tau}(\Pi_k) &= \mathcal{A}_*^{-1}(\Gamma_{\geq \tau}(\Pi_k^{\text{rect}})) \\ &= (\phi \circ \mathcal{A}_*)^{-1}(\Gamma_{\geq \tau}(\Pi_k^{\text{rect}})) \end{aligned}$$

from which it follows that

$$\text{vol}(\Gamma_{\geq \tau}(\Pi_k^{\text{rect}})) = \text{vol}(\Gamma_{\geq \tau}(\Pi_k)).$$

This conclusion holds for all  $\tau > 0$ , so it follows from the definition of the symmetrically decreasing rearrangement that  $(\Pi_k^{\text{rect}})^* \equiv \Pi_k^*$ .  $\square$

*Proof of Lemma 6.* We prove (1) by contradiction. Suppose that  $\Phi(\mathbf{t}) \geq \Psi(\mathbf{t})$  for all  $\mathbf{t}$ , but suppose  $\Phi^*(\mathbf{t}_0) < \Psi^*(\mathbf{t}_0)$  for some  $\mathbf{t}_0$ . Let  $\tau_0 := \Psi^*(\mathbf{t}_0)$ . Then

$$\text{vol}(\Gamma_{\geq \tau_0}(\Phi)) < \|\mathbf{t}_0\|^m v_m \leq \text{vol}(\Gamma_{\geq \tau_0}(\Psi)),$$

where  $v_m$  is the volume of the unit ball in  $\mathbb{R}^m$ . It follows that  $\Gamma_{\geq \tau_0}(\Psi) \setminus \Gamma_{\geq \tau_0}(\Phi)$  has positive measure, contradicting our assumption that  $\Phi(\mathbf{t}) \geq \Psi(\mathbf{t})$  for all  $\mathbf{t}$ .

Statement (2) follows from (1) by the observation that  $\max\{\Psi^*(\mathbf{t}), \tau\}$  is the symmetrically decreasing rearrangement of  $\max\{\Psi(\mathbf{t}), \tau\}$ .  $\square$

*Proof of Lemma 7.* Recall that

$$c_i := \max \left\{ \frac{1}{\gamma^2} \ln \left[ 1 + \alpha_i^\vee \left( 1 - \cos \frac{\gamma}{\sqrt{\alpha_i^\vee}} \right) \right], \quad \frac{1}{\alpha_i^\vee \pi^2} \ln \left[ 1 + 2\alpha_i^\vee \right] \right\}.$$

In particular,

$$c_i = \frac{1}{\gamma^2} \ln \left[ 1 + \alpha_i^\vee \left( 1 - \cos \frac{\gamma}{\sqrt{\alpha_i^\vee}} \right) \right] \quad \text{if } \alpha_i^\vee \geq \frac{\gamma^2}{\pi^2},$$

and

$$c_i = \frac{1}{\alpha_i^\vee \pi^2} \ln \left[ 1 + 2\alpha_i^\vee \right] \quad \text{if } \alpha_i^\vee \leq \frac{\gamma^2}{\pi^2}.$$

Define  $t_0 := \min \left\{ \frac{\gamma}{\sqrt{\alpha_i^\vee}}, \pi \right\}$ , and define  $f(t) := 1 + \alpha_i^\vee(1 - \cos t) - e^{c_i \alpha_i^\vee t^2}$  for  $-t_0 \leq t \leq t_0$ .

Note that  $f(0) = 0$ . Also, we claim that  $f(t_0) = 0$ . This must be verified in two cases, according to whether  $\alpha_i^\vee \geq \frac{\gamma^2}{\pi^2}$  or  $\alpha_i^\vee \leq \frac{\gamma^2}{\pi^2}$ .

If  $\alpha_i^\vee \geq \frac{\gamma^2}{\pi^2}$ , then  $t_0 = \frac{\gamma}{\sqrt{\alpha_i^\vee}}$ , so

$$\begin{aligned} f(t_0) &= 1 + \alpha_i^\vee \left( 1 - \cos \frac{\gamma}{\sqrt{\alpha_i^\vee}} \right) - \exp \left( \frac{\alpha_i^\vee}{\gamma^2} \cdot \ln \left[ 1 + \alpha_i^\vee \left( 1 - \cos \frac{\gamma}{\sqrt{\alpha_i^\vee}} \right) \right] \cdot \frac{\gamma^2}{\alpha_i^\vee} \right) \\ &= 0. \end{aligned}$$

If  $\alpha_i^\vee \leq \frac{\gamma^2}{\pi^2}$ , then  $t_0 = \pi$ , and

$$f(t_0) = 1 + 2\alpha_i^\vee - \exp \left( \frac{1}{\alpha_i^\vee \pi^2} \cdot \ln \left[ 1 + 2\alpha_i^\vee \right] \cdot \alpha_i^\vee \pi^2 \right) = 0.$$

This proves the claim that  $f(t_0) = 0$ . It follows that the average value of  $f'(t)$  on  $[0, t_0]$  is zero.

Finally, we observe that  $f'(0) = 0$ , and that  $f(t)$  has nonpositive third derivative on  $[0, t_0]$  (indeed, on  $[0, \pi]$ ). The verification of these claims is routine and is omitted. We infer that either  $f'(t) \equiv 0$  on  $[0, t_0]$ , or  $f''(t)$  has exactly one sign change on  $[0, t_0]$ , from positive to negative. In the latter case,  $f'(t)$  must also have exactly one sign change on  $[0, t_0]$  (also from positive to negative), since its average value on the interval is zero. It follows in either case that  $f(t) \geq 0$  on  $[0, t_0]$ , and thus on  $[-t_0, t_0]$  (since  $f(t)$  is an even function). This proves the lemma.  $\square$

Lemma 7 is used to establish Lemma 8.

*Proof of Lemma 8.* Let

$$K := \left\{ \mathbf{t} \in \mathbb{R}^m : |\mathbf{t}_i| \leq \min \left\{ \frac{\gamma}{\sqrt{\alpha_i^\vee}}, \pi \right\} \text{ for } i = 1, 2, \dots, m \right\}.$$

If  $\mathbf{t} \in K$ , then by Lemma 7,

$$\begin{aligned} \Pi_k^{\text{rect}}(\mathbf{t}) &= \prod_{i=1}^m \frac{1}{\sqrt{1 + \alpha_{(k-1)m+i}^\vee (1 - \cos \mathbf{t}_i)}} \\ &\leq \prod_{i=1}^m \frac{1}{\sqrt{1 + \alpha_i^\vee (1 - \cos \mathbf{t}_i)}} \\ &\leq \prod_{i=1}^m e^{-c_i \alpha_i^\vee \mathbf{t}_i^2 / 2}. \end{aligned}$$

Now suppose  $\mathbf{t} \notin K$ . Thus, there exists some  $i$  such that  $\mathbf{t}_i > \min \left\{ \frac{\gamma}{\sqrt{\alpha_i^\vee}}, \pi \right\}$ .

If  $\mathbf{t}_i > \pi$ , then we trivially have  $\Pi_k^{\text{rect}}(\mathbf{t}) = 0 \leq C'$ .

Otherwise, we have  $\mathbf{t}_i > \frac{\gamma}{\sqrt{\alpha_i^\vee}}$ , and therefore

$$\begin{aligned} \Pi_k^{\text{rect}}(\mathbf{t}) &\leq \frac{1}{\sqrt{1 + \alpha_i^\vee(1 - \cos \mathbf{t}_i)}} \\ &\leq \frac{1}{\sqrt{1 + \alpha_i^\vee(1 - \cos(\gamma/\sqrt{\alpha_i^\vee}))}} \\ &= e^{-\gamma^2 c_i/2} \\ &\leq C'. \end{aligned}$$

Thus whether  $\mathbf{t} \in K$  or  $\mathbf{t} \notin K$ , we have

$$\Pi_k^{\text{rect}}(\mathbf{t}) \leq \max \left\{ \prod_{i=1}^m e^{-c_i \alpha_i^\vee \mathbf{t}_i^2/2}, \quad C' \right\},$$

proving the lemma.  $\square$

**5.3. Upper bounds on  $C$ ,  $C'$ .** We now obtain

**Theorem 2b.** *Defining all constants as in the statement of Theorem 2a,*

$$C \leq \left[ \frac{\gamma}{2\sqrt{\pi \ln(1 + \frac{2\gamma^2}{\pi^2})}} \right]^m \prod_{i=1}^m \frac{1 - q_i^\vee}{\sqrt{q_i^\vee}}$$

and

$$C' \leq \frac{1}{\sqrt{1 + \frac{2\gamma^2}{\pi^2}}}.$$

*Remarks.* Notice that as  $\gamma \rightarrow \infty$ , all other inputs being fixed, we have  $C = O\left(\left(\frac{\gamma}{\ln \gamma}\right)^m\right)$  and  $C' = O\left(\frac{1}{\gamma}\right)$ . There is thus a trade-off between optimizing the  $Cp^{-m/2}$  term in Theorem 2a and optimizing the  $(C')^p$  term; the optimal choice of  $\gamma$  depends upon the other inputs.

Notice, also, that for fixed  $\gamma$  and for values of  $q_i^\vee$  bounded away from zero, the constant  $C$  is essentially a constant multiple of the bound on  $\text{conc}(AX)$  in Theorem 1. In fact, for (say)  $\gamma = 1$ , we have

$$C \leq (.657)^m \prod_{i=1}^m \frac{1 - q_i^\vee}{\sqrt{q_i^\vee}},$$

suggesting that the results of Theorem 2a are significantly better than those of Theorem 1 when  $p$  is large enough that the  $Cp^{-m/2}$  term dominates.



*Proof of Theorem 2b.* Recall that

$$c_i = \frac{1}{\gamma^2} \ln \left[ 1 + \alpha_i^\vee \left( 1 - \cos \frac{\gamma}{\sqrt{\alpha_i^\vee}} \right) \right] \quad \text{if } \alpha_i^\vee \geq \frac{\gamma^2}{\pi^2},$$

$$c_i = \frac{1}{\alpha_i^\vee \pi^2} \ln \left[ 1 + 2\alpha_i^\vee \right] \quad \text{if } \alpha_i^\vee \leq \frac{\gamma^2}{\pi^2},$$

$$C := \prod_{i=1}^m (2\pi c_i \alpha_i^\vee)^{-1/2},$$

and

$$C' := \max_{1 \leq i \leq m} e^{-\gamma^2 c_i / 2}.$$

Regarding  $c_i$  as a function of  $\alpha_i^\vee$ , we claim that this function is minimized at  $\alpha_i^\vee = \frac{\gamma^2}{\pi^2}$ . To demonstrate this claim, it suffices to check that:

- (1) The function  $f(x) := \frac{\ln(1+2x)}{x}$  is decreasing for  $0 < x \leq \frac{\gamma^2}{\pi^2}$ .
- (2) The function  $g(x) := x(1 - \cos \frac{\gamma}{\sqrt{x}})$  is increasing for  $\frac{\gamma^2}{\pi^2} \leq x < \infty$ .

*Proof of (1):* Differentiating, we obtain  $f'(x) = \frac{1}{x^2} \left[ \frac{2x}{1+2x} - \ln(1+2x) \right]$ . In general,  $\ln(1+u) > \frac{u}{1+u}$  for  $u > 0$ , so we have  $f'(x) < 0$  for all  $x > 0$ . In particular,  $f(x)$  is decreasing for  $0 < x \leq \frac{\gamma^2}{\pi^2}$ .

*Proof of (2):* Differentiating, we obtain  $g'(x) = 1 - \cos \frac{\gamma}{\sqrt{x}} - \frac{\gamma}{2\sqrt{x}} \sin \frac{\gamma}{\sqrt{x}}$ . It will be convenient to define  $y := y(x) = \frac{\gamma}{\sqrt{x}}$ . This change of variable bijectively transforms the interval  $\frac{\gamma^2}{\pi^2} \leq x < \infty$  into the interval  $0 < y \leq \pi$ . We may hence write  $g'(x) = h(y)$ , where

$$h(y) := 1 - \cos y - \frac{y}{2} \sin y.$$

Differentiating twice with respect to  $y$ , we obtain

$$\frac{dh}{dy} = \frac{1}{2} \sin y - \frac{y}{2} \cos y \quad \text{and} \quad \frac{d^2h}{dy^2} = \frac{y}{2} \sin y.$$

In particular, note that  $h(0) = 0$ ,  $h'(0) = 0$ , and  $h''(y) > 0$  for  $0 < y < \pi$ . It follows that  $h(y) > 0$  for  $0 < y \leq \pi$ . Equivalently,  $g'(x) > 0$  (and  $g(x)$  is increasing) for  $\frac{\gamma^2}{\pi^2} \leq x < \infty$ .

We have thus proved that  $c_i$  is minimized when  $\alpha_i^\vee = \frac{\gamma^2}{\pi^2}$ , in which case  $c_i = \frac{1}{\gamma^2} \ln \left( 1 + \frac{2\gamma^2}{\pi^2} \right)$ . That is to say,

$$c_i \geq \frac{1}{\gamma^2} \ln \left( 1 + \frac{2\gamma^2}{\pi^2} \right)$$

for all values of  $\alpha_i^\vee$ . It follows that

$$\begin{aligned} C &= \prod_{i=1}^m (2\pi c_i \alpha_i^\vee)^{-1/2} \leq \prod_{i=1}^m \left( \frac{2\pi}{\gamma^2} \ln \left( 1 + \frac{2\gamma^2}{\pi^2} \right) \cdot \frac{2q_i^\vee}{(1 - q_i^\vee)^2} \right)^{-1/2} \\ &= \left[ \frac{\gamma}{2\sqrt{\pi \ln \left( 1 + \frac{2\gamma^2}{\pi^2} \right)}} \right]^m \prod_{i=1}^m \frac{1 - q_i^\vee}{\sqrt{q_i^\vee}} \end{aligned}$$

and

$$\begin{aligned} C' &= \max_{1 \leq i \leq m} e^{-\gamma^2 c_i / 2} \leq \exp \left( - \frac{\ln \left( 1 + \frac{2\gamma^2}{\pi^2} \right)}{2} \right) \\ &= \frac{1}{\sqrt{1 + \frac{2\gamma^2}{\pi^2}}}, \end{aligned}$$

proving Theorem 2b. ■

## 6. PROOF OF THEOREM 3

We obtain Theorem 3 as a corollary of Proposition 3a, a more general result to follow. In order to state and prove Proposition 3a, we borrow the following notions from the theory of partially ordered sets (posets).

**Definitions 4.** Let  $S$  be a poset and  $x, y \in S$ . We say that  $x$  covers  $y$  if  $x > y$  and  $x \geq z \geq y \Rightarrow z \in \{x, y\}$ .

A rank function on a finite poset  $S$  is a function  $\text{rk} : S \rightarrow \mathbb{Z}_{\geq 0}$ , such that for all  $x, y \in S$ , if  $x$  covers  $y$ , then  $\text{rk}(x) = \text{rk}(y) + 1$ . We say that  $\text{rk}(x)$  is the rank of element  $x$ . A layer of a ranked poset is a level set of the rank function.

The chain of cardinality  $N$  is denoted by  $[N]$ , and is automatically assigned herein the unique rank function which assigns its least element rank 0. The product of two ranked posets  $S, S'$  is automatically assigned rank function equal to the sum of the rank functions of  $S, S'$ .

An antichain in a poset is a collection of pairwise incomparable elements. The width of a poset  $S$ , denoted by  $w(S)$ , is the cardinality of its largest antichain(s). The Whitney number  $W_i$  of a ranked poset is the cardinality of its layer of rank  $i$ . If the width of a ranked poset is equal to its largest Whitney number, then we say that the poset has the Sperner property.

For example, the “Boolean cube”  $([2] \times [2] \times [2])$  has Whitney numbers 1, 3, 3, 1 and width 3. Note that the width of any poset is greater than or equal to its largest Whitney number, because all layers are necessarily antichains.

Now we are ready to state

**Proposition 3a.** *Let  $X_1, X_2, \dots, X_p$  be independent, integer-valued random variables such that*

$$\text{conc}(X_j) \leq \frac{1}{N_j} \quad \text{for } 1 \leq j \leq p,$$

*where  $N_1, N_2, \dots, N_p$  are positive integers. Then*

$$\text{conc}(X_1 + \dots + X_p) \leq \frac{w([N_1] \times \dots \times [N_p])}{N_1 N_2 \dots N_p}.$$

*Moreover, given any fixed  $N$  such that  $2 \leq N_1, N_2, \dots, N_p < N$ , we have*

$$\frac{w([N_1] \times \dots \times [N_p])}{N_1 N_2 \dots N_p} \sim \left( \frac{\pi}{6} \sum_{j=1}^p (N_j^2 - 1) \right)^{-1/2}$$

*as  $p \rightarrow \infty$ .*

This proposition will be easiest to prove under the assumption that each  $X_j$  is uniformly supported on  $N_j$  points (with mass  $\frac{1}{N_j}$  at each). To justify passing to this case, we will use the following definition, and the two lemmas after it:

**Definition 5.** *A discrete random variable  $Y$  is a mixture of random variables  $Y_1, Y_2, \dots$  if its probability mass function lies in the convex hull of the probability mass functions of  $Y_1, Y_2, \dots$ .*

**Lemma 9.** *Let  $Y$  be a random variable, supported on  $\mathbb{Z}_{\geq 0}$ , such that  $\text{conc}(Y) \leq \frac{1}{N}$ . Then  $Y$  can be written as a mixture of random variables  $Y_1, Y_2, \dots$ , such that each  $Y_k$  is uniformly supported on  $N$  points, i.e., has an  $N$ -point support with probability mass  $\frac{1}{N}$  at each point in its support.*

*Proof of Lemma 9.* Let  $\mathcal{M}$  be the space of probability measures on  $\mathbb{Z}_{\geq 0}$ . Let

$$\mathcal{M}(N) := \left\{ \mu \in \mathcal{M} : \max_k \mu(\{k\}) \leq \frac{1}{N} \right\}$$

and

$$\mathcal{M}_u(N) := \{ \mu \in \mathcal{M} : \mu \text{ is uniformly supported on } N \text{ points} \}.$$

By the Krein-Milman theorem,  $\mathcal{M}(N)$  is the convex hull of its extreme points. We claim that the extreme points are precisely the points of  $\mathcal{M}_u(N)$ . It is immediately evident that each point of  $\mathcal{M}_u(N)$  is an extreme point of  $\mathcal{M}(N)$ . Conversely, suppose  $\mu \in \mathcal{M}(N) \setminus \mathcal{M}_u(N)$ . Thus there is some  $k \in \mathbb{Z}_{\geq 0}$  such that  $0 < \mu(\{k\}) < \frac{1}{N}$ , but in fact, there must be at least two distinct such  $k$ , since the total mass of  $\mu$  is 1 (an integer multiple of  $\frac{1}{N}$ ). Therefore,  $\mu$  is not an extreme point of  $\mathcal{M}(N)$ .

This proves our claim. Hence the probability measure associated to  $Y$  can be written as a countable convex combination of points of  $\mathcal{M}_u(N)$ , each of which defines the distribution of a random variable  $Y_k$  (proving the lemma).  $\square$

**Lemma 10** (Properties of superpositions). *If  $Y$  is a mixture of random variables  $Y_1, Y_2, \dots$ , then:*

- (1) *There is some  $k \geq 1$  for which  $\text{conc}(Y) \leq \text{conc}(Y_k)$ .*
- (2) *If  $Z$  is a random variable and  $f$  a function such that  $Z = f(Y)$ , then  $Z$  is a mixture of random variables  $Z_1, Z_2, \dots$ , where  $Z_k = f(Y_k)$ .*

*Proof of Lemma 10.* By the definition of mixture, there exist nonnegative  $\alpha_1, \alpha_2, \dots$  such that  $\alpha_1 + \alpha_2 + \dots = 1$  and such that

$$\Pr[Y = y] = \sum_{k=1}^{\infty} \alpha_k \Pr[Y_k = y].$$

Thus by the pigeonhole principle, for arbitrary  $y$ , there exists  $k = k(y)$  such that

$$\Pr[Y = y] \leq \Pr[Y_k = y].$$

Choosing  $y$  such that  $\text{conc}(Y) = \Pr[Y = y]$ , we conclude that  $\text{conc}(Y) \leq \text{conc}(Y_k)$  for this  $k$ . This proves claim (1) in the lemma. Claim (2) is self-evident.  $\square$

The heart of the proof of Proposition 3a is the following version of the local limit theorem:

**Definition 6.** *A sequence  $(\dots, b_{-1}, b_0, b_1, b_2, \dots)$  of nonnegative real numbers is properly log-concave if it is log-concave (i.e.,  $b_{t-1}b_{t+1} \leq b_t^2$  for all  $t$ ) and has no internal zeroes (i.e., if  $b_t > 0$  and  $b_{t+k} > 0$ , then  $b_{t+1}, b_{t+2}, \dots, b_{t+k-1} > 0$ ).*

**Lemma 11** (Bender). *Suppose that  $(\zeta_p : p \in \mathbb{N})$  is a sequence of integer-valued random variables,  $(F_p)$  are the corresponding distribution functions, and  $(\sigma_p)$  and  $(\mu_p)$  are sequences of real numbers such that  $\lim_{p \rightarrow \infty} F_p(\sigma_p x + \mu_p) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$  for every  $x \in \mathbb{R}$ . Also suppose that  $\sigma_p \rightarrow \infty$  as  $p \rightarrow \infty$ . Further, suppose that, for every  $p$ , the sequence  $b_p(t) := \Pr(\zeta_p = t)$  is properly log-concave with respect to  $t$ . Then*

$$\lim_{p \rightarrow \infty} \sigma_p \Pr(\zeta_p = \lfloor \sigma_p x + \mu_p \rfloor) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

*uniformly for all  $x \in \mathbb{R}$ .*

This result originally appeared in [2], but the above statement is based on its treatment in [8]; see either source for a proof.

*Proof of Proposition 3a.* For  $j = 1, 2, \dots, p$ , we are given to assume that  $\text{conc}(X_j) \leq \frac{1}{N_j}$ . By Lemma 9, each  $X_j$  is a superposition of some random variables which are each uniformly supported on some  $N_j$  points. Thus the random vector  $X = (X_1, \dots, X_p)$  is a mixture of random vectors each of the form  $X^{(k)} := (X_1^{(k)}, \dots, X_p^{(k)})$ , where the coordinates are independent and each  $X_j^{(k)}$  is uniformly supported on  $N_j$  points. The sum  $X_1 + \dots + X_p$  is a function of  $X$ , so by using both parts of Lemma 10, we see that

$$\text{conc}(X_1 + \dots + X_p) \leq \text{conc}(X_1^{(k)} + \dots + X_p^{(k)})$$

for some  $k$ . Since we are seeking an upper bound on  $\text{conc}(X_1 + \dots + X_p)$ , we assume with no loss of generality that  $X = X^{(k)}$ , or, more to the point, that each coordinate  $X_j$  is uniformly supported on  $N_j$  points (with mass  $\frac{1}{N_j}$  on each).

Denote the support of  $X_j$  by  $\{a_{j1}, a_{j2}, \dots, a_{jN_j}\}$ , where  $a_{j1} < a_{j2} < \dots < a_{jN_j}$ . Then

$$a_{1i_1} + a_{2i_2} + \dots + a_{pi_p} = a_{1i'_1} + a_{2i'_2} + \dots + a_{pi'_p}$$

implies that the  $p$ -tuples  $(i_1, i_2, \dots, i_p)$  and  $(i'_1, i'_2, \dots, i'_p)$  are identical or incomparable in  $[N_1] \times \dots \times [N_p]$ . It follows that

$$\text{conc}(X_1 + \dots + X_p) \leq \frac{w([N_1] \times \dots \times [N_p])}{N_1 N_2 \dots N_p}.$$

This proves the first claim of Proposition 3a.

For the remainder of the proof, assume that  $2 \leq N_1, N_2, \dots, N_p < N$  for some integer  $N$ . We are going to apply Lemma 13. Let  $\zeta_p$  denote the rank of a uniformly distributed random element of  $[N_1] \times [N_2] \times \dots \times [N_p]$ . Set  $\mu_p := \frac{N_1 + \dots + N_p}{2}$  and  $\sigma_p^2 = \sum_{j=1}^p \frac{N_j^2 - 1}{12}$ . It is easily verified that  $\mu_p$  and  $\sigma_p^2$  are respectively the mean and the variance of  $\zeta_p$ . By Lyapunov's central limit theorem [3], the condition

$$\lim_{p \rightarrow \infty} F_p(\sigma_p x + \mu_p) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

in Lemma 11 is satisfied. The hypothesis  $\sigma_p \rightarrow \infty$  is plainly also satisfied.

To see that the sequence  $b_p(t) := \mathbf{Pr}(\zeta_p = t)$  is properly log-concave, we note that this sequence is proportional to the Whitney numbers of the chain product  $[N_1] \times [N_2] \times \dots \times [N_p]$ , which is the convolution of the sequences of Whitney numbers for the factor chains. Each factor chain has Whitney numbers  $1, 1, \dots, 1, 0, 0, \dots$  (a properly log-concave sequence). Furthermore, the convolution of properly log-concave sequences is again properly log-concave, see e.g. [14]. Thus,  $(b_p(t))$  is properly log-concave.

All conditions of Lemma 11 have been verified, so the conclusion holds:

$$\lim_{p \rightarrow \infty} \sigma_p \mathbf{Pr}(\zeta_p = \lfloor \sigma_p x + \mu_p \rfloor) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

uniformly for all  $x \in \mathbb{R}$ . Setting  $x = 0$ , we obtain

$$\begin{aligned} \mathbf{Pr}(\zeta_p = \lfloor \mu_p \rfloor) &\sim \frac{1}{\sqrt{2\pi\sigma_p}} \\ &= \left( \frac{\pi}{6} \sum_{j=1}^p (N_j^2 - 1) \right)^{-1/2}. \end{aligned}$$

Finally, we observe that chain products have the Sperner property [8]. In particular, the width in the above formula is equal to the Whitney number  $W_{\lfloor \mu_p \rfloor}$ , so that

$$\frac{w([N_1] \times \cdots \times [N_p])}{N_1 N_2 \cdots N_p} = \mathbf{Pr}(\zeta_p = \lfloor \mu_p \rfloor).$$

This completes the proof of the proposition. ■

As an instance of Proposition 3a, we derive Theorem 3:

*Proof of Theorem 3.* As noted in the proof of Theorem 1, we have  $\text{conc}(X_j \mathbf{a}_j) = \frac{1}{\mathbf{E}(X_j)+1} \leq \frac{1}{\lfloor \mathbf{E}(X_j)+1 \rfloor}$  for  $1 \leq j \leq n$ . Since  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$  are linearly independent, we have

$$\begin{aligned} \text{conc}(AX) &= \prod_{i=1}^m \text{conc}(X_i \mathbf{a}_i + X_{m+i} \mathbf{a}_i + X_{2m+i} \mathbf{a}_i \cdots + X_{(p-1)m+i} \mathbf{a}_i) \\ &= \prod_{i=1}^m \text{conc}(X_i + X_{m+i} + X_{2m+i} + \cdots + X_{(p-1)m+i}) \\ &\lesssim \prod_{i=1}^m \left( \frac{\pi p}{6} (\lfloor \mathbf{E}(X_i) + 1 \rfloor^2 - 1) \right)^{-1/2}, \end{aligned}$$

where the last claim follows by Proposition 3a. Finally, by section 2.1, (2), we infer Theorem 3. ■

## 7. ACKNOWLEDGMENTS.

The author thanks Alexander Barvinok and Roman Vershynin for fruitful discussions, and the former also for making many helpful comments on the style of this paper.

## REFERENCES

- [1] A. Barvinok and J. Hartigan. Maximum entropy Gaussian approximation for the number of integer points and volumes of polytopes. *Advances in Applied Mathematics*, 45:252–289, 2010.
- [2] E. A. Bender. Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory, Series A*, 15:91–111, 1973.
- [3] P. Billingsley. *Probability and Measure, 2nd ed.* Wiley, New York, 1986.
- [4] A. Burchard. A short course on rearrangement inequalities. <http://www.math.utoronto.ca/almut/rearrange.pdf>.
- [5] J. De Loera. The many aspects of counting lattice points in polytopes. *Mathematische Semesterberichte*, 52:175–195, 2005.
- [6] J. De Loera. Counting and estimating lattice points: tools from algebra, analysis, convexity, and probability. *Optima: Newsletter of the Mathematical Programming Society*, Dec. 2009.
- [7] P. Diaconis and B. Efron. Testing for independence in a two-way table: new interpretations of the chi-square statistic. *Annals of Statistics*, 13:845–874, 1985.
- [8] K. Engel. *Sperner Theory*. Cambridge University Press, Cambridge, 1997.

- [9] P. Erdős. On a lemma of Littlewood and Offord. *Bulletin of the American Mathematical Society*, 51:898–902, 1945.
- [10] M.R. Garey and S.J. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, San Francisco, 1979.
- [11] G. Halász. Estimates for the concentration function of combinatorial number theory and probability. *Periodica Mathematica Hungarica*, 8:197–211, 1977.
- [12] R. Howard. Estimates on the concentration function of sets in  $\mathbb{R}^d$ : Notes on lectures of Oskolkov. <http://www.math.sc.edu/~howard/Notes/concentration.pdf>.
- [13] M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM*, 51:671–697, July 2004.
- [14] W. Kook. On the product of log-concave polynomials. *INTEGERS: Electronic Journal of Combinatorial Number Theory*, 6, 2006.
- [15] M. Rudelson and R. Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Advances in Mathematics*, 218:600–633, 2008.
- [16] T. Tao and V. Vu. Inverse littlewood-offord theorems and the condition number of random discrete matrices. *Annals of Mathematics*, 169:595–632, 2009.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109-1043, USA